ELSEVIER

PII: S0309-1708(96)00016-4

Advances in Water Resources, Vol. 19, No. 6, pp. 369–378, 1996

0309-1708/96/$15.00 + 0.00

Geostatistical interpolation of chemical concentration

Peter K. Kitanidis & Kuo-Fen Shen

Civil Engineering, Stanford University, Stanford, California 94305-4020, USA

(Received 19 February 1996)

Measurements of contaminant concentration at a hazardous waste site typically vary over many orders of magnitude and have highly skewed distributions. This work presents a practical methodology for the estimation of solute concentration contour maps and volume averages (needed for mass calculations) from data obtained from the analysis of water and soil samples. The methodology, which is an extension of linear geostatistics, produces a point estimate, i.e., a representative value, as well as a confidence interval, which contains the true value with a given probability. The approach uses a parsimonious model that accounts for the skewness by adding only one parameter to those used in linear geostatistics (variograms or generalized covariances). The resulting nonlinear kriging method is not substantially more difficult to use than linear geostatistics. The methodology is most appropriate when concentration measurements are available on a reasonably dense grid and no additional information (based on modeling flow and transport) can be used. We present and illustrate through an application, a practical approach to estimate all the parameters needed and to select and test the model. Copyright © 1996 Elsevier Science Ltd

Key words: geostatistics, best linear unbiased estimation, kriging, parameter estimation, maximum likelihood, restricted maximum likelihood, transformations, non-Gaussian, solute concentration.

INTRODUCTION

In the management of hazardous waste sites, one often has to estimate from available measurements the spatial extent of contaminant plumes or the total mass of chemicals (see Semprini *et al.*[1]). This information is essential in monitoring the progress of remediation for technical or legal purposes; it is also important in selecting the location and capacity of pumping wells in hydraulic containment and pump-and-treat systems or in designing enhanced *in situ* bioremediation projects. The most common method of measuring concentration is through the laboratory analysis of water and soil samples obtained in monitoring wells and soil borings, although soil-gas analysis and geophysical exploration techniques are sometimes used to complement the database. This work focuses on the geostatistical analysis of water- and soil-sample data. These measurements, which are typically a few dozen in number, vary over orders of magnitude and are arranged nonuniformly in space.

For illustration, Fig. 1 shows the histogram of data of trichloroethylene (TCE) in a vertical cross-section. One can see that the data are highly skewed. Even if one disregards analytical difficulties in the laboratory and uncertainties in the soil-water partition coefficients, the spatial variability in the observations makes it impossible to infer the exact location of the plume or the precise weight of the contaminants. In most cases, a cursory look at the data should make it obvious that one must provide an error bar that describes the reliability of any estimate.

Univariate statistical methods of data analysis, which treat the data as independent, are not applicable because data vary according to their relative location in space. For example, to compute the total mass of a substance, it would not be reasonable to assign equal weights to all concentration measurements because it is common to have more measurements near the center of the plume than elsewhere; instead, one should assign weights to measurements which are representative of the area of influence of each measurement. Linear geostatistical methods (also known as Best Linear Unbiased Estimation or BLUE methods[2,3]) account for spatial variability and are practical tools that have been applied in many

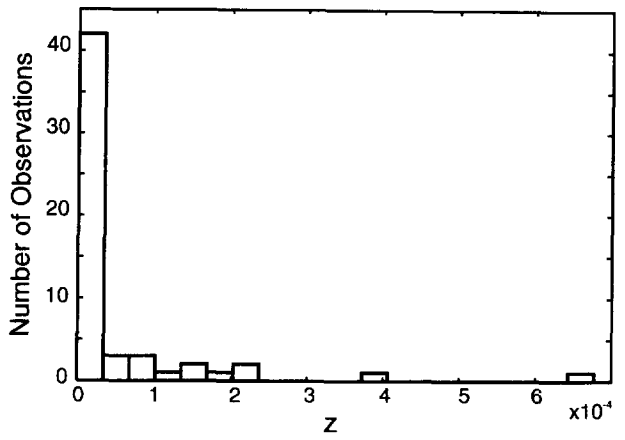

Fig. 1. Histogram of TCE concentration (μmol/liter) in a transect.

fields. They include the popular ordinary and universal kriging methods and are in some important ways similar to the prevalent methods of applied statistics.[4–7]

At first appearance, BLUE methods are applicable independently of distributional considerations because their derivation does not seem to assume any particular probability distribution. However, each step in the derivation of the BLUE methods makes better sense when some implicit conditions are met. Consider each step:

1. "The estimator is a linear function of the measurements." But such an estimator may be a poor choice for highly skewed, multimodal, or heavy-tailed distributions.
2. "The estimation error (which is modeled in estimation theory as a random variable) must have zero mean." However, if the estimation error has a severely skewed or bimodal distribution, its mean is not necessarily a satisfactory representative value and making the mean vanish may be less reasonable than, say, requiring that the median vanish.
3. "The estimation error is required to have as small a variance as possible." However, for an arbitrary distribution, the variance is not the sole measure of spread so that attempting to reduce the variance does not necessarily assure small errors. If the distribution of the errors is highly skewed, it may be far more relevant to applications to develop estimators that clip the tail of the error distribution, in order to reduce the chance of a very large error.

If it is somehow given that the probabilities are described through the multi-Gaussian distribution, the conditional mean is indeed a linear function of the data so that it is mathematically well established that in this special case the BLUE method is "optimal." If another distribution is given, one can in principle derive an estimator that is generally better in the variance or some other feature than the BLUE one.

Of course, in practical statistical modeling (as opposed to theoretic probabilistic analysis), it is data and not the distribution that is "given." For a BLUE method to make sense, the error distribution must not deviate dramatically from the Gaussian (normal) one. Then, requiring that the mean error be zero on the average and with as small a variance as possible assures minimization of the error in a generally acceptable sense. The cycle of the application of a BLUE method starts by assuming that the error is approximately Gaussian distributed and must close by testing this assumption against the experimental data. This is achieved by investigating the normality of residuals (i.e., whether the differences between observations and predictions follow a Gaussian distribution) through appropriate statistical tests, as discussed in Kitanidis.[8] This approach is universally accepted in applications of best linear unbiased estimation theory throughout statistics (e.g., see Belsley *et al.*,[9] p. 18; Draper and Smith,[6] Chapter 3 and references therein).

It is emphasized that the validity of the approach in geostatistics is not based on the metaphysical question of whether the sampled process is truly multi-Gaussian. It is not necessary, not possible, and perhaps not meaningful to demonstrate from a finite sample that the field is indeed multi-Gaussian. Nevertheless, it is customary in statistics (where only low order statistics, i.e., first two moments are employed) to use the multi-Gaussian model as a guide to develop effective parameter estimation methods and tests.[4–7] The results are usually insensitive to moderate deviations from the multi-Gaussian assumption and can be supported using intuitive least-squares arguments. For example, the restricted maximum likelihood method (which can be used to estimate variograms) can be interpreted as modified least-squares fitting or "cross-validation."[8]

In the analysis of concentration data the problem is that estimation errors may vary over orders of magnitude, depending on whether the estimated concentration is at a "hot spot," and are highly skewed. Linear estimation methods (such as ordinary kriging) do not perform well because the distribution of estimation errors is not described adequately by the mean and the mean square value. One particular disturbing feature is that the concentration 95% confidence interval may include negative values. This leaves for consideration nonlinear geostatistical methods such as disjunctive, indicator, and probability kriging. Although in theory these methods have significant potential, in practice their applicability is limited because:

1. In the authors' opinion, practicable methods for parameter estimation and model testing for these methods have yet to be developed. Thus, the practitioner has little guidance in evaluating the validity of the model or in choosing the right parameters.

2. In effect, they involve a large number of options or "parameters" that are hard to select from the data.
3. They are computationally intensive making it difficult to test a model or perform sensitivity analysis and to compare different models.

The objective of this work is to present an alternative nonlinear estimation method that has the following advantages:

1. Parsimony, i.e., it uses empirical models with few parameters, a major consideration in any estimation method.
2. Availability of practical methods for parameter estimation and model testing.
3. Computational efficiency. Following a transformation, the data are analyzed using linear geostatistical methods.
4. To some extent, applicability using available geostatistical codes.

In terms of organization, the paper consists of two parts: the first part presents the theoretical basis of the approach and the second part describes how the methodology is used in practice and presents some examples.

THE MODEL

Transformation

Consider a positive spatial process $z(\mathbf{x})$ (i.e., $z(\mathbf{x}) > 0$) and the transformation

$$y(\mathbf{x}) = \begin{cases} \dfrac{z(\mathbf{x})^{\kappa} - 1}{\kappa}, & \text{if } \kappa \neq 0 \\ \ln[z(\mathbf{x})], & \text{if } \kappa = 0 \end{cases} \tag{1}$$

where κ is a parameter. This transformation is frequently used in statistics.[6,10,11] Transformations have also been used in geostatistics (e.g., Verly[12]) with emphasis on the logarithmic transformation. This paper develops a new method for determining the transformation parameter and applies it to a wider class (not just ordinary kriging) than other works.

The basic premise is that $y(\mathbf{x})$ may be modeled as a multi-Gaussian process. For example, if $z(\mathbf{x})$ is the concentration at location $\mathbf{x}$ and $\kappa = 0$, then the logarithm of the concentration, $y(\mathbf{x})$, is modeled as multi-Gaussian and we say that the concentration $z(\mathbf{x})$ is log-Gaussian distributed. The log-Gaussian distribution is quite useful in applications and has been the subject of several studies in the context of estimation of spatial functions (e.g., Switzer and Parker[13]). Note that transformation (1), which is known as the *power transformation*, is more general because it includes the untransformed case ($\kappa = 1$), the square root transformation ($\kappa = \frac{1}{2}$), etc.

In essence, our model is that the experimental $z(\mathbf{x})$ is a realization of a random field with a multivariate probability distribution that includes as special cases the Gaussian, the log-Gaussian, and other distributions. The parameters of this model are:

(1) the parameters of a geostatistical model for the transformed data;
(2) the transformation parameter κ.

For a given κ, we can proceed to transform the data and to fit a geostatistical model to the transformed data. However, how do we judge which κ fits the data? The fit of the variogram to the transformed data is not a reliable criterion. For instance, for concentration data, the logarithmic transformation may produce "better behaving" data (because it suppresses more of the variability of the original data) than the less drastic square root transformation. However, because we want to predict concentrations, we must at the end back-transform from y to z, which means computing the exponent if $\kappa = 0$ or computing the square if $\kappa = \frac{1}{2}$. The best estimate and confidence interval of z depend much more on the higher moments of y if we obtain z from y through exponentiation than through squaring. Because of inherent uncertainties about higher moments, exponentiation is less desirable than squaring.

Thus, we should select the transformation parameter and the variogram of the transformed data jointly in order *to fit a model to the concentration measurements* rather than to their transformation. A way to achieve this goal is to apply maximum likelihood estimation methods to determine the value of the parameter κ.[6,10]

Geostatistical model

The probability distribution of process z is defined from the distribution of process y through the transformation of eqn (1). We will assume initially that the mean function and the covariance function of y are known expressions involving some parameters which we will estimate from data; but we will soon see how to relax these assumptions as is the geostatistical practice (i.e., if $y(\mathbf{x})$ is intrinsic, we will only need to estimate the parameters of its variogram).

We will facilitate the derivation by using a compact vector notation that is common in statistics. Let:

$\mathbf{y} = n$ by 1 vector of transformed data, i.e., $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix}$

$\mathbf{m} = n$ by 1 mean of $\mathbf{y}$, i.e., the ith element of $\mathbf{m}$ is the expected value of y_i; $\mathbf{Q} = n$ by n covariance matrix of $\mathbf{y}$,

i.e., the ijth element of $\mathbf{Q}$ is the covariance between y_i and y_j;

$\mathbf{z} = n$ by 1 vector of original data, i.e., $\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \cdots \\ z_n \end{bmatrix}$

In linear geostatistics, the mean is a linear function of the drift coefficients. Thus,

$$\mathbf{m} = \mathbf{X}\boldsymbol{\beta} \tag{2}$$

where $\mathbf{X}$ is an $n \times p$ matrix of known coefficients and $\boldsymbol{\beta}$ is the $p \times 1$ vector of the coefficients of the mean, also known as *drift coefficients* (see Kitanidis[14] for a more detailed discussion of this model in the context of geostatistics). The covariance matrix is a function of some parameters $\boldsymbol{\theta}$,

$$\mathbf{Q} = \mathbf{Q}(\boldsymbol{\theta}) \tag{3}$$

For illustration, consider the intrinsic case with exponential covariance function: $p = 1$, β = the constant mean of the intrinsic function, and

$$\mathbf{X} = \begin{bmatrix} 1 \\ 1 \\ \cdots \\ 1 \end{bmatrix} \Bigg\} n$$

The ijth element of $\mathbf{Q}$ is $\theta_1 \exp(-|\mathbf{x}_i - \mathbf{x}_j|/\theta_2)$, where θ_1 is the variance and θ_2 is the correlation length.

ML PARAMETER ESTIMATION

The probability density function $p(\mathbf{y})$ of $\mathbf{y}$ is Gaussian with mean $\mathbf{m}$ and covariance $\mathbf{Q}$,

$$p(\mathbf{y}) = (2\pi)^{-n/2}\det[\mathbf{Q}]^{-1/2}\exp(-\tfrac{1}{2}(\mathbf{y}-\mathbf{m})^T\mathbf{Q}^{-1}(\mathbf{y}-\mathbf{m})) \tag{4}$$

where det[ ] indicates determinant, $^{-1}$ inverse, and ^{T} transpose of a matrix.

The pdf $p(\mathbf{z})$ of the original data is

$$p(\mathbf{z}) = p(\mathbf{y})\left|\frac{d\mathbf{y}}{d\mathbf{z}}\right| \tag{5}$$

where $|d\mathbf{y}/d\mathbf{z}|$ is the absolute value of the Jacobian determinant of the data transformation. Since:

$$y_i = \begin{cases} \dfrac{z(\mathbf{x}_i)^\kappa - 1}{\kappa}, & \text{if } \kappa \neq 0 \\ \ln[z(\mathbf{x}_i)], & \text{if } \kappa = 0 \end{cases}$$

then

$$\left|\frac{d\mathbf{y}}{d\mathbf{z}}\right| = \prod_{i=1}^{n} z(\mathbf{x}_i)^{\kappa-1} \tag{6}$$

Thus, the distribution of the original data $z(\mathbf{x})$ for any n sampling locations $z(\mathbf{x}_1), z(\mathbf{x}_2), \ldots, z(\mathbf{x}_n)$ is

$$p(\mathbf{z}) = (2\pi)^{-n/2}\det[\mathbf{Q}]^{-1/2} \times \exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{m})^T\mathbf{Q}^{-1}(\mathbf{y}-\mathbf{m})\right)\prod_{i=1}^{n} z(\mathbf{x}_i)^{\kappa-1} \tag{7}$$

The parameters of the model are κ and the parameters of the mean and the covariance function of the process $y(\mathbf{x})$. If these parameters are unknown but $z(\mathbf{x}_1)$, $z(\mathbf{x}_2), \ldots z(\mathbf{x}_n)$ are available, then eqn (7) represents the likelihood of the data and is proportional to the probability distribution of the parameters given the data. In the *maximum likelihood* (ML) method, estimates of the parameters are obtained by maximizing the likelihood function. In practice, it is more convenient to minimize minus the logarithm of the likelihood function:

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \kappa) = \frac{n}{2}\ln(2\pi) + \frac{1}{2}\ln(\det[\mathbf{Q}]) + \frac{1}{2}(\mathbf{y}-\mathbf{m})^T\mathbf{Q}^{-1}(\mathbf{y}-\mathbf{m}) - \ln\left(\prod_{i=1}^{n} z(\mathbf{x}_i)^{\kappa-1}\right) \tag{8}$$

This can be written as the sum of two functions:

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}, \kappa) = L_g(\boldsymbol{\beta}, \boldsymbol{\theta}|\kappa) + L_p(\kappa) \tag{9}$$

where

$$L_g(\boldsymbol{\beta}, \boldsymbol{\theta}|\kappa) = \frac{n}{2}\ln(2\pi) + \frac{1}{2}\ln(\det[\mathbf{Q}]) + \frac{1}{2}(\mathbf{y}-\mathbf{m})^T\mathbf{Q}^{-1}(\mathbf{y}-\mathbf{m}) \tag{10}$$

$$L_p(\kappa) = -\ln\left(\prod_{i=1}^{n} z(\mathbf{x}_i)^{\kappa-1}\right) \tag{11}$$

Thus, parameter estimation in the ML approach can proceed as follows:

1. For a given value of κ, transform the data and find the parameters β and θ which minimize L_g. This can be achieved using methods described in Kitanidis and Lane.[15] Additionally, compute L_p.
2. Repeat the procedure with another value of κ.
3. Select the value of κ that minimizes the sum $L_g + L_p$.

RML PARAMETER ESTIMATION

In ordinary and universal kriging, the values of the drift parameters are not required because unbiasedness constraints are introduced to make the estimator independent of these coefficients. Furthermore, instead of the ordinary covariance function, the process may be characterized by the generalized covariance function which can be estimated from the data without

knowledge of the drift coefficients.[3,16] For example, intrinsic functions employed in ordinary kriging are characterized by the variogram, which is the negative of a generalized covariance function. The point is that, in practice, one is interested in estimating the parameters of a generalized covariance function.

As discussed in detail in Kitanidis,[17] the parameters of the generalized covariance function can be obtained by maximizing the marginal distribution of the covariance function parameters conditional on the data, i.e., the distribution obtained after averaging over all possible values of the drift coefficients. For a given value of κ, the marginal probability density function of $\boldsymbol{\theta}$ can be obtained from the joint probability density function of $(\boldsymbol{\beta}, \boldsymbol{\theta})$ through integration over $\boldsymbol{\beta}$. Thus, making use of eqn (2):

$$\int_{\beta}(2\pi)^{-n/2}\det[\mathbf{Q}]^{-1/2}\exp\left(-\tfrac{1}{2}(\mathbf{y}-\mathbf{m})^T\mathbf{Q}^{-1}(\mathbf{y}-\mathbf{m})\right)d\boldsymbol{\beta}$$
$$= (2\pi)^{-(n-p)/2}\det[\mathbf{Q}]^{-1/2}\det[\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X}]^{-1/2}$$
$$\times\exp\left(-\tfrac{1}{2}\mathbf{y}^T(\mathbf{Q}^{-1}-\mathbf{Q}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Q}^{-1})\mathbf{y}\right) \tag{12}$$

This method is known as *restricted maximum likelihood* (RML). The same result can be obtained as follows: transform the data into a set of $n-p$ authorized increments and write down its likelihood function. Then maximize the value of the likelihood function with respect to the parameters of the covariance function. The RML method has deep roots in statistical inference, in relation to the fundamental problem of eliminating unwanted parameters (see, e.g., discussion in Edwards[18]). Patterson and Thompson[19] introduced such a method in the context of analysis of variance problems and Kitanidis[20] introduced RML in the inference of random fields and showed for the first time that RML is a general method for the estimation of generalized covariance functions used by Matheron.[3] For the latter problem, the unwanted parameters are the drift coefficients and the problem is how to make inferences without dealing with values for these parameters.

The expression for minus the logarithm of the restricted loglikelihood of the data is:

$$R(\boldsymbol{\theta},\kappa) = R_g(\boldsymbol{\theta}|\kappa) + L_p(\kappa) \tag{13}$$

where

$$R_g(\boldsymbol{\theta}|\kappa) = \frac{n-p}{2}\ln(2\pi) + \frac{1}{2}\ln(\det[\mathbf{Q}])$$
$$+\tfrac{1}{2}\ln(\det[\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X}])$$
$$+\frac{1}{2}\mathbf{y}^T(\mathbf{Q}^{-1}-\mathbf{Q}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{Q}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Q}^{-1})\mathbf{y} \tag{14}$$

$$L_p(\kappa) = -\ln\left(\prod_{i=1}^{n} z(\mathbf{x}_i)^{\kappa-1}\right) \tag{15}$$

Thus, parameter estimation in the RML approach can proceed as follows:

1. For a given value of κ transform the data and find the parameters $\boldsymbol{\theta}$ which minimize R_g. Additionally, compute L_p.
2. Select the value of κ which optimizes the sum $R_g + L_p$.

ESTIMATION

Once the appropriate transformation parameter has been computed and the geostatistical model of the transformed data has been selected, it is a straightforward task to infer the transformed variable using a BLUE method, such as ordinary kriging. Kriging provides at each point a best estimate, $\hat{y}$, and a mean square estimation error, σ^2. Under the assumption that the transformed variable is multi-Gaussian, the best estimate is the conditional mean and the mean square error is the conditional variance.

However, in most applications, we are expected to present the results in terms of concentrations (the original variable) and not the transformed variable. It is straightforward to write down the probability distribution of the concentration conditional on the observations and the modeling assumptions outlined above.

$$p(z) = (2\pi\sigma^2)^{-1/2}\exp\left(-\frac{(\ln(z)-\hat{y})^2}{2\sigma^2}\right)z^{-1}, \text{ if } \kappa = 0 \tag{16}$$

$$p(z) = (2\pi\sigma^2)^{-1/2}\exp\left(-\frac{\left(\frac{z^\kappa-1}{\kappa}-\hat{y}\right)^2}{2\sigma^2}\right)z^{\kappa-1},$$
$$\text{if } \kappa \neq 0 \tag{17}$$

This distribution is not Gaussian and not even symmetric, except of course in the trivial case that $\kappa = 1$. One difficulty with non-symmetric distributions is that there is no single number that can be used as a representative value because the median, the mode, and the mean differ from each other. They are given below.

Median. The median is the value that is exceeded 50% of the time. When one makes a one-to-one transformation, the median transforms to the median. Thus,

$$\text{med}(z) = \begin{cases}(\kappa\hat{y}+1)^{\frac{1}{\kappa}}, & \text{if } \kappa \neq 0\\ \exp[\hat{y}], & \text{if } \kappa = 0\end{cases} \tag{18}$$

Mode. The mode is the most likely value. That is, it corresponds to the point where the probability

distribution has a maximum.

$$\text{mod}(z) = \begin{cases} \left(\dfrac{1+\kappa\hat{y}+\sqrt{1+4\kappa\sigma^2(\kappa-1)+\kappa\hat{y}(2+\kappa\hat{y})}}{2}\right)^{\frac{1}{\kappa}}, & \text{if } \kappa \neq 0 \\ \exp[\hat{y}-\sigma^2], & \text{if } \kappa = 0 \end{cases} \tag{19}$$

Mean. The mean is

$$\text{mean}(z) = \begin{cases} m(\hat{y},\sigma^2,\kappa) & \text{if } \kappa \neq 0 \\ \exp[\hat{y}+\sigma^2/2], & \text{if } \kappa = 0 \end{cases} \tag{20}$$

where $m(\hat{y},\sigma^2,\kappa)$ is a function, presented in a series form in the Appendix, that depends on $\hat{y}$, σ^2, and κ.

Typically, concentration measurements are positively skewed so that $\text{mod}(z) \leq \text{med}(z) \leq \text{mean}(z)$. See, e.g., Fig. 2 that shows $p(z)$ and the mode, median, and mean for $\hat{y} = 4$, $\sigma^2 = 1$, $\kappa = 0$. Of those three, the median is the most straightforward to compute and is more stable (i.e., less affected by sampling error and more resistant to extremes) than the mean in the case of skewed distributions. For this reason, in many applications, the median is preferable as a "point" or "best" estimate. The median is unbiased in the sense that the error has equal chances of being positive or negative. However, in the expected value sense, the median tends to underestimate the true value, since the median is less than the mean.

In linear estimation, the accuracy of the estimate is usually given through the mean square error or the standard estimation error (also known as root mean square error). It is feasible to calculate the mean square error:

$$\text{var}(z) = \begin{cases} v(\hat{y},\sigma^2,\kappa) & \text{if } \kappa \neq 0 \\ \exp[2\hat{y}+\sigma^2](\exp[\sigma^2]-1), & \text{if } \kappa = 0 \end{cases} \tag{21}$$

where $v(\hat{y},\sigma^2,\kappa)$ in series form is given in the Appendix.

The mean square error is useful when the distribution of the errors is nearly Gaussian so that the mean square error describes adequately the distribution of the errors about the representative value. However, for skewed distributions, such as the one shown in Fig. 2, there is no single measure of spread that is universally satisfactory. In exploratory analysis, the interquartile range (or "Q-spread") is used. The interquartile range is the difference between the 0·75 quantile (i.e., the value that is not exceeded 75% of the time) and the 0·25 quantile.

The most straightforward approach is to compute the 0·25 and 0·75 quantiles and also the interquartile range. This is accomplished as follows:

$$\begin{aligned} \hat{y}_{0\cdot75} &= \hat{y} + 0\cdot675\sigma \\ \hat{y}_{0\cdot25} &= \hat{y} - 0\cdot675\sigma \end{aligned} \tag{22}$$

Then back-transform to find the $\hat{z}_{0\cdot75}$ and the $\hat{z}_{0\cdot25}$ values and finally compute the interquartile range

$$I_q = \hat{z}_{0\cdot75} - \hat{z}_{0\cdot25} \tag{23}$$

The result of concentration estimation is then presented graphically as shown in Fig. 3. This plot shows the best estimate, the range of values that contain the concentration with probability 50%, and also shows the asymmetry in the estimation errors.

APPLICATION

We will illustrate the approach using data from a sand aquifer where the groundwater is contaminated with trichloroethylene (TCE) and its products from natural anaerobic reductive dechlorination.[1] Water samples were collected on five vertical transects at multiple wells and levels (depths) and analyzed at the laboratory

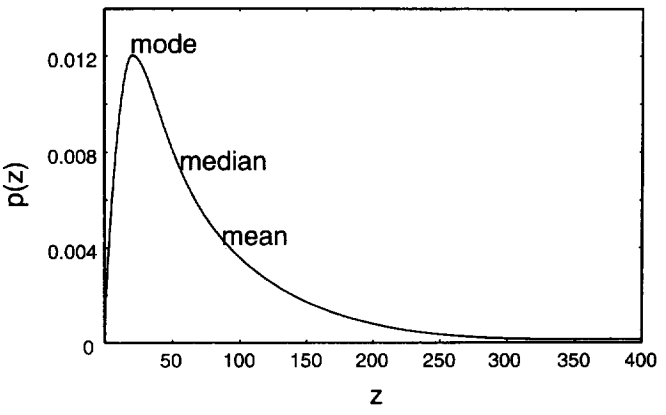

Fig. 2. Mode, median, and mean of lognormal distribution.

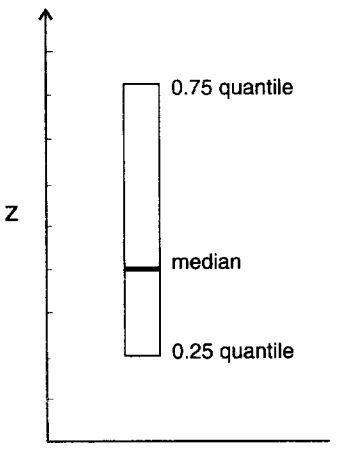

Fig. 3. Representative value and error bar of concentration estimate.

to measure the concentration of some organic and inorganic compounds. All these concentration data were then analyzed using the techniques presented in this work and contour maps were prepared depicting the distribution of the mass of each compound over each cross-section. The objective of this study was to illuminate how TCE degrades under natural conditions and particularly to study the effect of redox conditions. Here, for illustrative purposes, we will summarize the results of the analysis of the TCE data from transect 1, a data set which contains 58 measurements.

Before starting, let us review the three basic steps involved in the development of an empirical model:

(1) *Exploratory analysis* is where data and other information are used to select the type of models to be considered. Then a model with a few adjustable parameters is tentatively chosen.
(2) *Parameter estimation* means obtaining from the data good estimates of the parameters conditional on the adequacy of the assumed model.
(3) *Validation or diagnostic checking* is defined as "checking the fitted model in its relation to the data with intent to reveal model inadequacies and so to achieve model improvement" (Box and Jenkins,[7] p. 171). Diagnostic checking may result in a new model, in which case the procedure must be repeated.

Attention was limited to the class of models described in this work, which includes as special cases ordinary and log-Gaussian kriging with isotropic or anisotropic covariance functions or variograms. The parameter estimation problem is thus reduced to selection of the transformation parameter κ and the variogram of the transformed variable. Diagnostic checking is performed through evaluation of the residuals.

For the TCE concentration data set, we performed the power transformation for different κ values, with κ increasing from 0 to 1 with increments of 0·1. The geostatistical model that we used for the transformed data is intrinsic isotropic with a linear variogram,

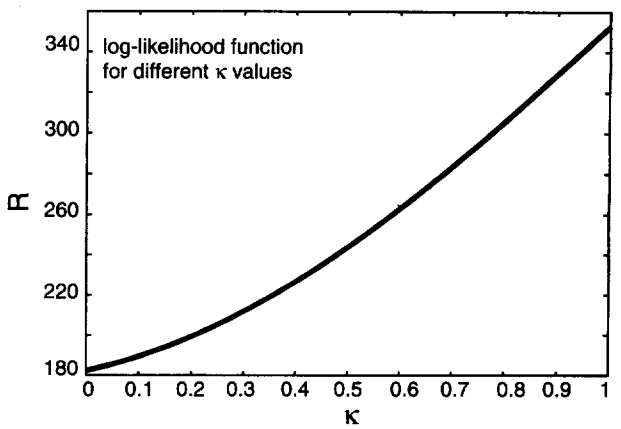

Fig. 4. Optimization of the value of the κ parameter.

Table 1. Parameter estimation of the transformed TCE data

α	θ_1	cR	R
1	0·2670	3·349	164·6
2	0·1249	2·78	159·3
4	0·06169	2·333	154·3
5	0·05172	2·289	153·8
6	0·04609	2·304	154·0
7	0·04268	2·356	154·3

$\gamma(h) = \theta_1 h$. Implementing the RML parameter estimation method, we found the slope θ_1 for each κ value and computed the sum $R_g + L_p$. The results, which are shown in Fig. 4, indicate that the value of κ that optimizes the sum $R = R_g + L_p$ in the interval [0, 1] is $\kappa = 0$. Thus, tentatively, the logarithmic transformation is selected for this case. The original TCE concentration data is extremely skewed, while the distribution of their logarithms is much closer to the Gaussian. Note, however, that since the log-concentration observations are correlated, one cannot perform one of the common goodness-of-fit tests, such as Filliben.[21] Instead, the important question is whether orthonormal residuals are Gaussian-like and this question can be answered through goodness-of-fit tests.

Next, we proceed with the linear geostatistical analysis of the transformed data. It appears that the correlation decays faster in the vertical direction than in the horizontal direction. To account for this anisotropy in the correlation structure, we introduced an additional parameter, α, that stretches the vertical coordinate, rendering the correlation structure isotropic (in the transformed domain). We focused on a variogram that is the superposition of the nugget and the linear but when we optimized we found that the nugget effect vanishes. From Table 1 we see that the best estimates are:

α (the stretching coefficient) = 5,
θ_1 (the slope of the variogram) = 0·05172.

In Table 1, cR is the geometric mean of the variances of the orthonormal residuals of the transformed residuals[8] and measures the goodness of fit of the model to the transformed data. We also confirmed that for the above value of the stretching coefficient, the optimal value of κ is still 0.

The exponential variogram model is also considered. The estimation of the model parameters yielded the mode: $\gamma(h) = 10(1 - \exp(-h/200))$. However, the fit (value of cR) is practically the same as with the simpler linear model. For this reason, we kept the simpler linear one.

Before using the linear model to make predictions, we tested the residuals of the process and compared with the ordinary kriging approach:

- The orthonormal residuals conformed to the Gaussian distribution at the 95% significance level. Their histogram is shown in Fig. 5.

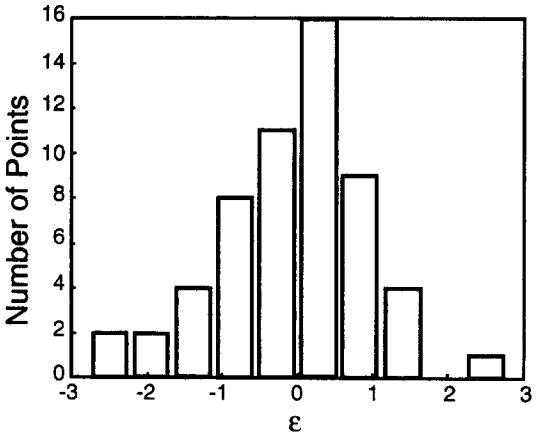

Fig. 5. Histogram of orthonormal residuals of TCE data.

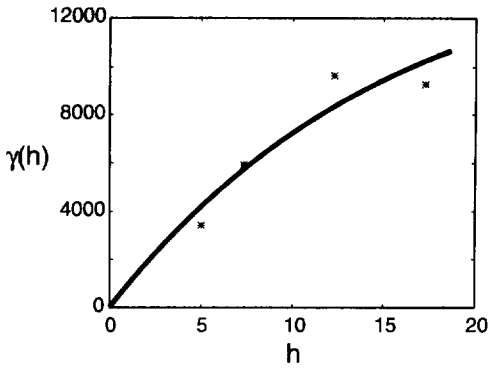

Fig. 6. Variogram of original TCE data.

- For comparison purposes, we used as base approach ordinary kriging with a variogram fitted graphically on the experimental variogram of the original data, as shown in Fig. 6. We then compared the fit of the two approaches as follows. We used the first observation to predict the second TCE observation, the first two observations to predict the third, and so on. The mean absolute error using our approach was 40% smaller than that of the base procedure. The suboptimality of

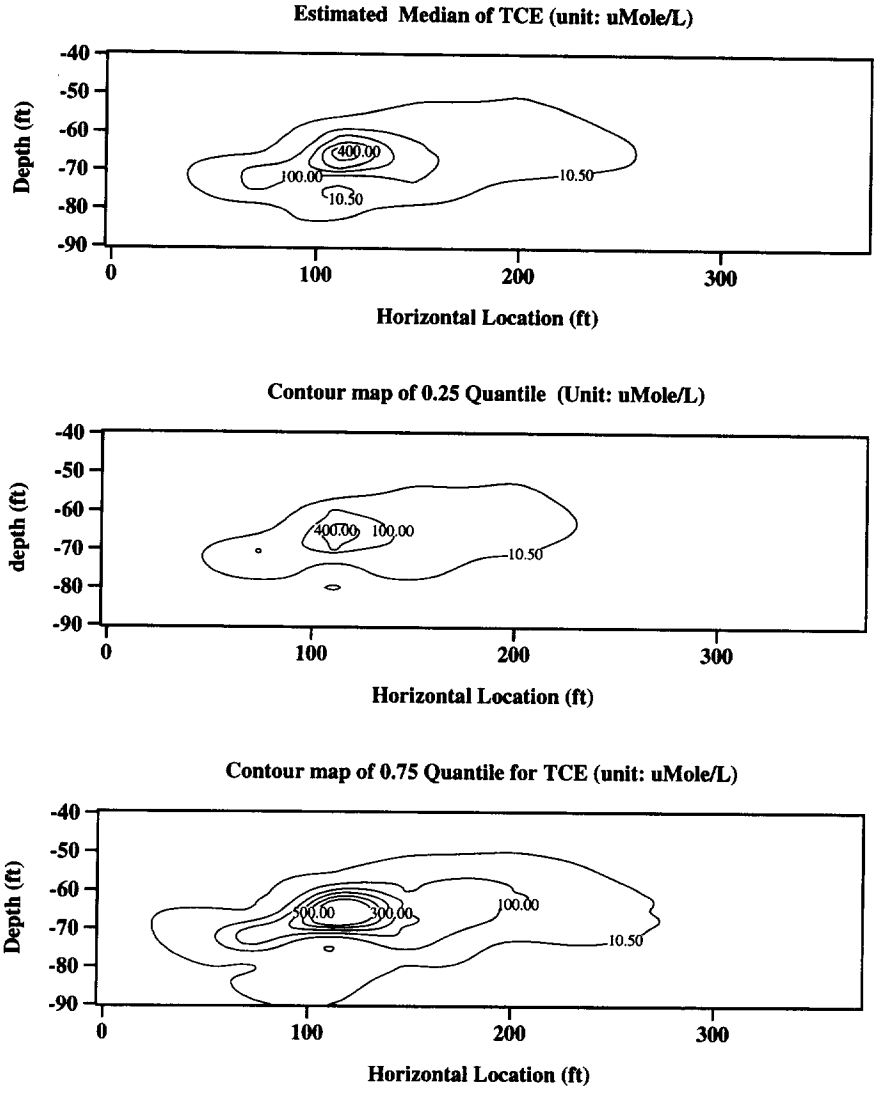

Fig. 7. Contour maps of the median, the lower quartile, and the upper quartile of TCE estimates.

the base approach is partially due to the fact that the residuals are severely non-Gaussian. The orthonormal residuals of the base approach fail the Filliben test for normality even at the 75% significance level.

Figure 7 shows the representative value (median) and the lower and upper quartiles of the TCE concentration as estimated from the linear model. The physical significance of the results is discussed in Semprini *et al.*,[1] where additional results from the application of this methodology are presented.

CONCLUDING REMARKS

BLUE and other linear interpolators are prevalent in solving interpolation problems. However, the direct application of BLUE methods in interpolating solute concentration measurements may lead to poor estimates and questionable error bars, when the distribution of data violates assumptions implicit in BLUE.

One way to improve estimation is by introducing additional information such as information embedded in models of flow, transport, and chemical transformation, as in Graham and McLaughlin.[22] This approach would be preferable in principle, because the structure used in interpolation of data would be based on physical considerations. However, in the study that motivated our work[1] and in many other cases encountered in practice, these physical and chemical processes are complex, inadequately understood, and involve many other parameters. If observations of concentration are sufficiently dense, a more practicable approach is to describe the structure through empirical models that are properly selected using statistical methods and are simple to apply in drawing isoconcentration lines.

We proposed here a relatively simple nonlinear interpolation method that is based on a convenient nonlinear transformation in combination with linear estimation methods. We discussed the logical underpinnings of the method, developed the tools for its implementation, showed its application to an actual case, and showed that the results compare favorably with the base case of interpolation through ordinary kriging.

ACKNOWLEDGMENTS

Funding for this study was provided by the Office of Research and Development, U.S. Environmental Protection Agency, under agreement R-815738-01 through the WRHSRC. Funding for this work has also been provided by the National Science Foundation under grant BCS-8914812. The content of this paper does not necessarily represent the views of these agencies.

REFERENCES

1. Semprini, L., Kitanidis, P. K., Kampbell, D. & Wilson, J. T., Anaerobic transformation of chlorinated aliphatic hydrocarbons in a sand aquifer based on spatial chemical distributions. *Water Resour. Res.*, **31(4)** (1995) 1051–1062.
2. Matheron, G., *The Theory of Regionalized Variables and its Applications*. Ecole des Mines, Fontainbleau, France, 1971.
3. Matheron, G., The intrinsic random function. *Adv. Appl. Prob.*, **5** (1973) 438–468.
4. Rao, C. R., *Linear Statistical Inference and its Applications*, Wiley, New York, 1973, 625 pp.
5. Schweppe, F. C., *Uncertain Dynamics Systems*. Prentice-Hall, Englewood Cliffs, NJ, 1973.
6. Draper, N. & Smith, H., *Applied Regression Analysis*. 2nd edition, Wiley-Interscience, New York, 1981.
7. Box, G. E. P. & Jenkins, G. M., *Time Series Analysis*. Holden-Day, San Francisco, 1976.
8. Kitanidis, P. K., Orthonormal residuals in geostatistics; Model criticism and parameter estimation. *Math. Geology*, **23** (1991) 741–758.
9. Belsley, D. A., Kuh, E. & Welsch, R. E., *Regression Diagnostics*. Wiley, NY, 1980.
10. Box, G. E. P. & Tiao, G. C., *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, MA, 1973.
11. Howarth, R. J. & Earle, S. A. M., Application of a generalized power transformation to geochemical data. *Math. Geology*, **11** (1979) 45–62.
12. Verly, G., The multi-Gaussian approach and its application to the estimation of local reserves. *Math. Geology*, **15** (1983) 259–286.
13. Switzer, P. & Parker, H. M., The problem of ore versus waste discrimination for individual blocks: The lognormal model. In *Advanced Geostatistics in the Mining Industry*, eds M. Guarascio, M. David & C. Huijbregts. Reidel, Dordrecht-Holland, 1976.
14. Kitanidis, P. K., Parametric estimation of covariances of regionalized variables. *Water Resour. Bull.*, **23** (1987) 557–567.
15. Kitanidis, P. K. & Lane, R. W., Maximum likelihood parameter estimation of hydrologic spatial processes by the Gauss–Newton method. *J. Hydrology*, **79** (1985) 53–71.
16. Kitanidis, P. K., Generalized covariance functions in estimation. *Math. Geology*, **25(5)** (1993) 525–540.
17. Kitanidis, P. K., Parameter uncertainty in estimation of spatial functions: Bayesian analysis. *Water Resour. Res.*, **22(4)** (1985) 499–507.
18. Edwards, A. W. F., *Likelihood*. The Johns Hopkins Univ. Press, 1992, 275 pp.
19. Patterson, H. D. & Thompson, R., Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58(3)** (1971) 545–554.
20. Kitanidis, P. K., Statistical estimation of polynomial generalized covariance functions and hydrologic applications. *Water Resour. Res.*, **19** (1983) 909–921.
21. Filliben, J. J., The probability plot correlation test for normality. *Technometrics*, **17(1)** (1975) 111–117.
22. Graham, W. D. & McLaughlin, D., Stochastic analysis of nonstationary subsurface solute transport; 1. Unconditional moments. *Water Resour. Res.*, **25(2)** (1989) 215–232.

APPENDIX

For the case $\kappa \neq 0$, the mean is

$$\text{mean}(z) = \int_{-\infty}^{\infty} (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{\left(\frac{z^\kappa - 1}{\kappa} - \hat{y}\right)^2}{2\sigma^2}\right] z^{\kappa-1} \cdot z\,dz$$

$$= \int_{-\infty}^{\infty} (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[\frac{-(y-\hat{y})^2}{2\sigma^2}\right] (y\kappa+1)^{\frac{1}{\kappa}} dy$$

Using the Taylor series expansion for $(y\kappa+1)^{\frac{1}{\kappa}}$ about the mean $\hat{y}$:

$$(y\kappa+1)^{\frac{1}{\kappa}} = (1+\kappa\hat{y})^{\frac{1}{\kappa}} + (1+\kappa\hat{y})^{\frac{1}{\kappa}-1}(y-\hat{y}) + \frac{(1-\kappa)}{2!}(1+\kappa\hat{y})^{\frac{1}{\kappa}-2}(y-\hat{y})^2 + \frac{(1-\kappa)(1-2\kappa)}{3!}(1+\kappa\hat{y})^{\frac{1}{\kappa}-3}(y-\hat{y})^3 + \cdots + \frac{(1-\kappa)(1-2\kappa)\cdots[1-(n-1)\kappa]}{n!} \times (1+\kappa\hat{y})^{\frac{1}{\kappa}-n}(y-\hat{y})^n.$$

then we substitute this into the equation for mean (z) and get

$$\text{mean}(z) = (1+\kappa\hat{y})^{\frac{1}{\kappa}} + 0 + \frac{(1-\kappa)}{2!}(1+\kappa\hat{y})^{\frac{1}{\kappa}-2}\sigma^2 + 0 + \frac{(1-\kappa)(1-2\kappa)(1-3\kappa)}{4!}(1+\kappa\hat{y})^{\frac{1}{\kappa}-4}3\cdot\sigma^4 + 0 + \begin{cases} 0 & \text{if } n = \text{odd} \\ \dfrac{(1-\kappa)(1-2\kappa)\cdots(1-(n-1)\kappa)}{n!}(1+\kappa\hat{y})^{\frac{1}{\kappa}-n}\cdot 1\cdot 3\cdot 5\cdots(n-1)\sigma^n & \text{if } n = \text{even} \end{cases}$$

As for the variance σ^2, we can use the relationship $\sigma_z^2 = E[z^2] - \text{mean}(z)^2$, and $E[z^2]$ can be calculated by the following equation:

$$E[z^2] = \int_{-\infty}^{\infty} (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{\left(\frac{z^\kappa - 1}{\kappa} - \hat{y}\right)^2}{2\sigma^2}\right] \cdot z^{\kappa-1} \cdot z^2 dz$$

$$= \int_{-\infty}^{\infty} (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{(y-\hat{y})^2}{2\sigma^2}\right] \cdot (y\kappa+1)^{\frac{2}{\kappa}} dy$$

Similarly, we use the Taylor series expansion for $(y\kappa+1)^{\frac{2}{\kappa}}$ about the mean $\hat{y}$:

$$(y\kappa+1)^{\frac{2}{\kappa}} = (1+\kappa\hat{y})^{\frac{2}{\kappa}} + 2\cdot(1+\kappa\hat{y})^{\frac{2}{\kappa}-1}(y-\hat{y}) + (2-\kappa)(1+\kappa\hat{y})^{\frac{2}{\kappa}-2}(y-\hat{y})^2 + \frac{(2-\kappa)(2-2\kappa)}{3}(1+\kappa\hat{y})^{\frac{2}{\kappa}-3}(y-\hat{y})^3 + \cdots + 2\cdot\frac{(2-\kappa)(2-2\kappa)\cdots[2-(n-1)\kappa]}{n!} \times (1+\kappa\hat{y})^{\frac{2}{\kappa}-n}(y-\hat{y})^n$$

Then $E[z^2]$ becomes:

$$(1+\kappa\hat{y})^{\frac{2}{\kappa}} + (2-\kappa)(1+\kappa\hat{y})^{\frac{2}{\kappa}-2}\sigma^2 + \frac{(2-\kappa)(2-2\kappa)(2-3\kappa)}{12}(1+\kappa\hat{y})^{\frac{2}{\kappa}-4}\cdot 3\cdot\sigma^4 + \cdots + \begin{cases} 0 & \text{if } n = \text{odd} \\ \dfrac{2}{n!}(2-\kappa)(2-2\kappa)\cdots[2-(n-1)\kappa] & \\ \quad\times(1+\kappa\hat{y})^{\frac{2}{\kappa}-n}\cdot 1\cdot 3\cdot 5\cdots(n-1)\sigma^n, & \text{if } n = \text{even} \end{cases}$$